Confirmatory Factor Analysis of the Patient Health Questionnaire-9: A Study of the Participants From the Spinal Cord Injury Model Systems

James S. Krause, PhD, Lee L. Saunders, PhD, Charles Bombardier, PhD, Claire Kalpakjian, PhD

Objective: To examine the factor structure of the Patient Health Questionnaire-9, a measure of depression, in persons with spinal cord injury (SCI).

Design: Cross-sectional, confirmatory factor analytic study.

Setting: Community.

Participants: Data for 7296 persons with an SCI who had sustained their injury at least 1 year prior to assessment and who had complete Patient Health Questionnaire-9 data collected at a follow-up interview were drawn from the National Spinal Cord Injury Statistical Center Database.

Interventions: None.

Main outcome measures: Factor structure of the Patient Health Questionnaire-9.

Results: Confirmatory factor analysis indicated a marginal fit for the single factor solution (root mean square error of approximation [RMSEA] = 0.086), whereas the solution with 5 somatic items and 4 nonsomatic items had the best fit (RMSEA = 0.054) among 2-factor models that used all 9 items. Of the models that used fewer than 9 items, the best fit was for the 6-item solution with 3 somatic items (sleep, appetite, and fatigue) and 3 nonsomatic items (feeling down, feeling bad about self, and suicidal ideation; RMSEA = 0.043). Similar results were found across the strata except for the Hispanic group (for whom no model fit well).

Conclusions: Given the results of this analysis that support a 2-factor structure of the Patient Health Questionnaire-9 in persons with SCI, the next step in this line of research is to validate each of these dimensions against other ways of measuring depression.

PM R 2011;3:533-540

INTRODUCTION

Depression is an important complication after spinal cord injury (SCI). Rates of major depression reported in the literature have ranged from 11% to more than 30% after SCI [1,2]. Differing rates may be reflective of the varying measurement techniques across studies. Recently a systematic review of depression measurement in SCI found very little psychometric evaluation of measurement tools in the past 27 years, despite the attention given to depression in this population [3].

Controversy remains regarding the validity of certain items included in depression screening measures. In the context of healthy persons and those without physical disabilities, somatic symptoms such as sleep disturbance, anergia, and appetite changes may represent manifestations of a depressive disorder. However, among people with medical conditions or disabilities, these somatic symptoms may represent secondary conditions or complications of the physical condition rather than an indication of a depressive disorder. This ambiguity concerning symptom etiology is not restricted to SCI but is present among medical patients, people with disabilities, the elderly, women, children and adolescents, culturally diverse groups, prison populations, and the poor [4]. Gastrointestinal symptoms, sleep disturbance, headaches, appetite changes, fatigue, and aches and pains of a diffuse nature are common features of depression, making their association with depression quite

J.S.K. Department of Health Sciences and Research, College of Health Professions, Medical University of South Carolina, 77 President St, C101, MSC 700, Charleston, SC 29425 Address correspondence to J.S.K.; e-mail: krause@musc.edu
Disclosure: 8, NIDRR (grants H133N060009, H133B090005)

L.L.S. College of Health Professions, Medical University of South Carolina, Charleston, SC
Disclosure: 8, NIDRR (grant H133B090005)

C.B. Department of Rehabilitation Medicine, University of Washington, Seattle, WA
Disclosure: 1A, Pfizer stock; 8, NIDRR (grants H133N060033, H133A060107)

C.K. Department of Physical Medicine and Rehabilitation, University of Michigan, Ann Arbor, MI
Disclosure: 8, NIDRR (grant H133N060032)

Peer reviewers and all others who control content have no relevant financial relationships to disclose.

The contents of this manuscript were developed under grants from the Department of Education, NIDRR grant numbers H133N060032, H133N060033, H133A060107, and H133N060009. However, those contents do not necessarily represent the policy of the Department of Education, and you should not assume endorsement by the Federal Government.

Disclosure Key can be found on the Table of Contents and at www.pmrjournal.org

Submitted for publication October 13, 2010; accepted March 2, 2011.

challenging to determine in the context of SCI, particularly for those newly injured or with other medical conditions.

This measurement question has important implications for clinical practice. If a valid depression screening measure is found to be unifactorial, then clinicians would have more confidence adopting the "inclusive approach" [5] to screen for major depression. That is, they could count all symptoms toward the probable diagnosis of depression without having to scrutinize the etiology of each symptom. Alternatively, if a depression screening measure is multifactorial, then clinicians must rule out the possibility that a symptom is related etiologically to a medical condition or physical disability before attributing it to a depressive disorder.

In previous research on the topic of depression screening in people with SCI, Bombardier et al [6] examined the positive predictive power of each item of the Patient Health Questionnaire-9 (PHQ-9) [7]. The authors reported that all of the items composing this measure were efficient predictors of the total score and that it was therefore appropriate to use the inclusive approach to screen for major depression in persons with SCI.

Another approach is to use factor analysis to identify the dimensionality or factor structure of an instrument. Some well-known depression screening measures, such as the Beck Depression Inventory II [8], are known to be multifactorial [9], including in medical samples [10]. However, a factor analytic study of the *Diagnostic and Statistical Manual of Mental Disorders III-revised* (DSM III-R) [11] criteria suggests that the symptoms make up a single factor [12]. In addition, factor analytic studies of the PHQ-9 have confirmed a single factor in a racially and ethnically diverse sample of mostly female primary care patients [13], people with vision loss [14], and people with substance use disorders [15]. Factor analysis of depression scales has been widely used to examine symptom coverage, to examine the degree to which items represent different symptom clusters, or to determine the equivalence of factor structures across subsamples (eg, age or gender). It also has been used to conduct comparison studies to determine treatment effectiveness for specific symptoms, the specificity of certain symptoms, or whether subgroups of patients can be distinguished on the basis of specific symptom factors [16].

Factor analysis may be particularly useful in understanding the relationship of the broad dimensions of depression symptom types, such as somatic, cognitive, and affective domains. The relationship among these factors may provide some insight into the role of somatic depressive symptoms in the context of SCI or other disabling conditions. A number of recent studies have used the PHQ-9 after SCI and have reported varying results regarding the factor structure. Richardson and Richards [17] used the National Spinal Cord Injury Statistical Center (NSCISC) database to assess the PHQ-9 factor structure and stratified by time since injury. They used exploratory factor analysis (EFA) with principal-axis factoring, Promax rotation, and loadings of 0.40 or greater in factor designation. Their results indicated that a 2-factor solution fit best across all years after injury, and 6 items consistently loaded on the same factors (somatic: sleep, appetite, fatigue; nonsomatic [NS]: feeling hopeless, feeling bad, suicidal ideation). The other 3 items (little interest, psychomotor changes, trouble concentrating) either loaded on each of the 2 factors, depending on time since injury, or did not load with either factor.

Kalpakjian et al [18] also applied EFA to the PHQ-9 with use of a subset of cases from the NSCISC database. Men and women were matched on level/completeness of SCI, follow-up year, and current age to examine the congruence of factor structure across gender. Specifically, 1- and 2-factor models and congruence between them for 2 randomly split half samples were examined with the use of principal-axis factor analyses and oblique (ie, Promax) rotation; this approach was used because factors were not presumed to be independent. Congruence of factors between genders was tested with the Pearson correlation coefficient and the Tucker congruence coefficient. Results indicated that 1- and 2-factor solutions fit the structure of the items, with original item variance ranging from 41% to 51%. Congruence between random samples was uniformly high for the 1-factor structure (r = 0.79-0.95) but variable for the 2-factor solution. Although it was high for the combined sample of women and men and for men only (r = 0.90-0.97 and 0.71-0.94, respectively), it was variable for the women only (r = 0.29-0.85). The authors suggest that the low congruence between genders and within women for the 2-factor structure may point to important differences about how certain symptoms may be experienced or interpreted differently by men and women with SCI.

With use of data from the NSCISC database, Graves and Bombardier [19] assessed the factor structure of the PHQ-9 with confirmatory factor analysis (CFA). Although they concluded there was only 1 factor in the PHQ-9, they only tested the unidimensional model and reported a root mean square error of approximation (RMSEA) of 0.091 (95% confidence interval = 0.086-0.097). This finding exceeds the standard of 0.080 or less for acceptable fit [20,21].

Most recently, Krause et al [22] applied CFA to a sample of 568 newly injured participants with SCI during their inpatient rehabilitation hospitalization by testing 4 competing models. In addition to the unidimensional model, each of 3 two-factor models included fatigue, appetite change, and sleep loss in a somatic factor. One of the models added the item "psychomotor changes" to the somatic factor, and another model dropped this item altogether (ie, 5 items versus 3). The solution with 3 somatic items and 6 NS items had an acceptable fit (RMSEA = 0.073), as did the 8-item solution that dropped the psychomotor changes item.

Table 1. *Description of each CFA model run*

Model	Factors	No. Items	Description
Model A	1	9	Little interest, feeling down, sleep, appetite, fatigue, feeling bad about self, trouble concentrating, psychomotor, suicidal ideation
Model B	2	9	Somatic (sleep, appetite, fatigue) Affective (little interest, feeling down, feeling bad about self, trouble concentrating, psychomotor, suicidal ideation)
Model C	2	9	Somatic (sleep, appetite, fatigue, psychomotor) Affective (little interest, feeling down, feeling bad about self, trouble concentrating, suicidal ideation)
Model D	2	9	Somatic (fatigue, sleep, appetite, trouble concentrating, psychomotor) Affective (feeling down, feeling bad about oneself, suicidal ideation, little interest)
Model E	2	8	Somatic (sleep, appetite, fatigue) Affective (little interest, feeling down, feeling bad about self, trouble concentrating, suicidal ideation)
Model F	2	8	Somatic (fatigue, sleep, appetite, little interest, trouble concentrating) Affective (suicidal ideation, feeling bad about self, feeling down)
Model G	2	6	Somatic (sleep, appetite, fatigue) Affective (feeling down, feeling bad about self, suicidal ideation)

CFA = confirmatory factor analysis.

The authors conducted a follow-up approximately 12 months after discharge [23]. The somatic factor measured during inpatient rehabilitation was not significantly predictive of either somatic or NS factors at follow-up. In contrast, the NS factor was significantly related to both somatic and NS factors. These findings question the utility of somatic items during inpatient rehabilitation in predicting future outcomes.

One major limitation in existing literature is the absence of studies on the generalizability of the factor structure of depressive symptoms across different racial-ethnic groups. Research in primary care suggested similar levels of internal consistency to the single scale model of the PHQ-9 across multiple racial-ethnic groups [13]. However, some differences were noted in item responses. Previous research with SCI has indicated that race-ethnicity is a factor related to differential reporting of depressive symptoms. For instance, nonwhite (primarily black) participants were more likely to report depressive symptoms and that these symptoms related to socioeconomic status [24]. Using a smaller but more diverse sample, Kemp et al [25] found that Hispanic participants reported greater overall depression scores than did white or black participants. These studies do not provide a basis for assuming similar factor structures across race-ethnicity among participants with SCI.

On the basis of studies of the DSM-III-R [11], we may expect the symptoms of major depression to be unifactorial. Recent research suggests that the PHQ-9, which is based on the DSM-IV [26] criteria for a major depressive episode, is multifactorial, but no clear consensus exists regarding factor structure of the PHQ-9 in persons with SCI. The majority of studies have used the national data set with variable results. Only 2 studies have used CFA, which is more powerful than EFA, but one of these studies only tested a single, 1-factor model [19]. The other study used participants during inpatient rehabilitation [22], and results may be taken as strong evidence for a 2-factor model during inpatient rehabilitation, but these findings clearly cannot be applied to participants living in the community.

In summary, the existing literature has provided several alternative factor models that use the PHQ-9 among those with SCI. However, with the exception of a single study in which the authors administered the measure during inpatient rehabilitation [22], no investigators have compared alternative models using CFA. Similarly, with the exception of a single study that matches cases on gender [27], no authors have compared the factor structure as a function of other demographic and injury characteristics, such as race-ethnicity or injury severity.

Purpose and Hypothesis

The purpose of this study was to perform CFA of 7 competing factor models (Table 1) to identify the factor structure of the PHQ-9 in persons at least 1 year after SCI onset. Each of the models tested was determined by the results of the study of inpatients [22] or identified from EFA studies on community samples [17,18].

This study includes 3 primary research hypotheses: (1) Each of the 2-factor models identified in previous research will have a better fit, as indicated by a lower RMSEA, than the 1-factor model; (2) the 2-factor models will have a better fit than the 1-factor model for each of the subsamples on the basis of gender, race-ethnicity, and injury severity; and (3) among the 2-factor models, the solution that uses only 6 items will provide the best fit because it essentially consists of the 3 best-fitting items under each factor.

METHODS

Participants

We included all persons in the NSCISC database from the SCI model systems (SCIMS), for which data collection began in 1973. SCIMS are selected on the basis of excellence in clinical care and research and are located throughout the United States. Participants are recruited during inpatient rehabilitation after injury, and information is collected during rehabilitation (form I). Thereafter follow-up is performed by phone, mail, or in person at 1 year, 5 years, and every fifth year after injury (form II). Although follow-up was attempted at those milestones, some participants who were difficult to reach at those times may have undergone follow-up in a different year (ie, instead of year 1, they underwent follow-up at year 2). At each SCIMS site, participants provided informed consent to participate in the study. The PHQ-9 was added to the follow-up data collection tool in October 2000. Our study included persons who participated in follow-up interviews from October 2000 through November 30, 2006, because the PHQ-9 was collected only during this time period. Only participants ages 18 years and older were asked the PHQ-9 questions. In addition, persons who made a full recovery from their injury (American Spinal Injury Association Impairment Scale [ASIA] E) were excluded from the analysis.

At least one assessment was completed for 10,978 unique individuals between 2000 and 2006, of which 7301 had information on the PHQ-9. For this study, we included persons who participated in at least 1 follow-up in which the PHQ-9 was assessed. For persons who completed more than 1 PHQ-9, the most recent was retained. From 7301 participants who had information on the PHQ and met inclusion criteria, 7296 had complete PHQ-9 data and were retained (5 persons made a complete recovery).

Measures

The PHQ-9 is a depression screening instrument validated in primary medical care settings [7]. The PHQ-9 consists of 9 items that parallel the symptoms that define DSM-IV [26] criteria of major depressive disorder (MDD; eg, having little interest or pleasure in doing things and feeling down, depressed, or hopeless). The respondent is asked to identify how frequently each symptom has been a problem during the past 2 weeks. The following scores are assigned for each of the category responses: (0) not at all; (1) several days; (2) more than half of the days; and (3) nearly every day. Using a simple additive scoring model and a cutoff of at least 10 to indicate MDD, Kroenke et al [7] reported good sensitivity (0.88) and specificity (0.88) compared with an independent structured diagnostic interview for major depression in a primary care sample. Kroenke et al [7] also reported good internal consistency (0.89) and test-retest reliability (0.84).

Other variables used in this analysis were gender, race (ie, white, black, and Hispanic) and injury severity. Injury severity was classified according to both function (ie, ASIA-A-C versus ASIA-D) and level of injury (ie, C1-C4, C5-C8, non-cervical).

Analysis

Mplus 5.0 [28], a specialized program for latent variable modeling, was used to test the factor structure of the PHQ-9 with the use of CFA with maximum likelihood estimation. Because the response categories of the PHQ-9 do not fully represent ratio scaling, we used the categorical command (a special feature of Mplus). The RMSEA was used to assess the fit of the model, with a value of <0.05 representing outstanding fit, <0.08 representing adequate fit, <0.10 representing mediocre fit, and >0.100 as unacceptable [29]. We also used the Comparative Fit Index (CFI) to assess model fit, with a value of 0.90 or greater representing acceptable fit [30]. Because all CFI were greater than the cutoff of 0.90, indicating a good fit, we did not report individual CFI values. None of the factors had fewer than 3 items [31].

We tested 7 CFA models developed on the basis of previous literature (Table 1). Model A is a 1-factor model that uses all 9 items [19]. Models B, C, and D are 2-factor models tested by Krause et al [22]; all have a somatic factor and an NS factor. Richardson and Richards [17] developed Models E and F with 2 factors by using 8 items with a somatic factor and an NS factor. Model G, also by Richardson and Richards, is a 2-factor model with somatic and NS factors but only includes 6 items total [17]. For each of the models, we performed CFA on the full sample and also for specific groups on gender, race-ethnicity, and injury severity.

RESULTS

Participant Characteristics

On average, participants were 31.8 years of age at injury (SD = 13.9) and 42.0 years of age at follow-up (SD = 13.7). In addition, 33.2% were in their first year of follow-up, 16.3% were between 2 and 5 years' follow-up, 13.1% were between 6 and 10 years' follow-up, and 37.4% were at more than 10 years' follow-up. Seventy-two percent of the sample were white, 17.6% were black, and 8.0% were Hispanic. Seventy-nine percent were men. Of participants, 52.2% had a cervical injury, 52.1% had an ASIA-A injury, 12.9% had an ASIA-B injury, 12.9% had an ASIA-C injury, and 22.1% had an ASIA-D injury. The average score on the PHQ-9 was 4.5 (SD = 5.5).

One-Factor Model

A 1-factor model was tested that contained all items of the PHQ-9 (Model A). The RMSEA was 0.086 for a 1-factor

Table 2. *Results from CFA using the RMSEA*

Strata	n	Model A	Model B	Model C	Model D	Model E	Model F	Model G
No. of factors		1	2	2	2	2	2	2
No. of items		9	9	9	9	8	8	6
No. of somatic items		—	3	4	5	3	5	3
No. of affective items		—	6	5	4	5	3	3
Correlation between factors			0.66	0.68	0.68	0.66	0.68	0.59
RMSEA								
Race								
White	5273	0.089	0.065*	0.068*	0.052*	0.064*	0.080*	0.042
Black	1283	0.077*	0.065*	0.066*	0.052*	0.064*	0.063*	0.039
Hispanic	582	0.094	0.074*	0.081	0.079*	0.078*	0.095	0.077*
Gender								
Male	5767	0.083	0.064*	0.065*	0.050†	0.063*	0.073*	0.039†
Female	1529	0.092	0.068*	0.074*	0.067*	0.066*	0.089	0.058*
Injury severity								
C1-C4, nonambulatory	990	0.082	0.061*	0.070*	0.053*	0.061*	0.070*	0.034†
C5-C8, nonambulatory	1691	0.093	0.065*	0.066*	0.057*	0.069*	0.088	0.057
Noncervical, nonambulatory	2867	0.079*	0.061*	0.063*	0.049†	0.061*	0.067*	0.035†
Ambulatory	1585	0.086	0.064*	0.066*	0.055*	0.061*	0.081	0.045†
Total sample	7296	0.086	0.064*	0.067*	0.054*	0.063*	0.076*	0.043†

CFA = confirmatory factor analysis; RMSEA = root mean square error of approximation.
*Please supply footnote.
†Please supply footnote.

solution when the entire sample was used, representing marginal fit (Table 2). The CFI indicated an acceptable fit of greater than 0.90 for the full sample, but the value (0.96) was lower than that of any other solution (greater scores indicate better fit).

Two-Factor Models With Nine Items

Of the 2-factor models that used all 9 items (models B, C, and D), model D, the solution with 5 somatic items (trouble sleeping, feeling tired, poor appetite, trouble concentrating, and psychomotor) and 4 NS items, had the lowest RMSEA (0.054 compared with 0.064 and 0.067).

Two-Factor Models With Fewer Than Nine Items

Of the solutions that used 8 or fewer items (models E, F, and G), by far the best fit was observed for the 6-item solution (Model G) with 3 somatic items (sleep, appetite, and fatigue) and 3 NS items (feeling down, feeling bad about self, and suicidal ideation). The RMSEA for the total sample was 0.043 and was the best fit of all models.

Factor Models Across Subgroups

Generally, similar results were found across the strata with respect to adequacy of fit of the models. For almost all strata, model G had the lowest RMSEA. The exception was for the Hispanic group, in whom model B had the lowest RMSEA (0.074). However, none of the models tested in the Hispanic group showed excellent fit.

DISCUSSION

Each of the 3 study hypotheses was confirmed. The results suggest that, even though depression is conceptualized as a univariate factor, the 1-factor model is not an acceptable fit, both for the sample as a whole and among strata on the basis of gender, race-ethnicity, and injury severity. In contrast, several 2-factor models produced an acceptable fit, with the solution using only 6 items producing the best fit. Thus in a large sample of persons with SCI, the PHQ-9 appears to be best represented by distinct somatic and NS items, although the 2 factors are correlated, indicating 2 distinct yet nonorthogonal dimensions. The 2-factor structure (with all or some of the somatic items) may suggest that, among those with complex health situations caused by disability or other health conditions, depression cannot be simply characterized by a combination of NS and somatic symptoms. Rather, a 2-dimensional understanding of the experience of depression, with perhaps dominance of NS symptoms (which would need to be further validated), may be a more refined way of understanding a complex disorder such as depression in the context of disability. Although the fit varied somewhat as a function of participant characteristics, with Hispanic participants having greater RMSEA, the same general pattern supporting the 2-factor model held across subgroups.

Clinical Implications

Depression diagnosis in people with comorbid medical conditions can be approached in at least 2 different ways. The "inclusive approach" dictates that all symptoms in a depression screening measure or structured interview contribute to

a diagnosis of depression and should be included in the scoring or diagnostic process [5]. This approach is considered more reliable because judgments do not have to be made about the etiology (attributable to the primary medical condition versus a psychiatric disorder) of a given symptom and may be more sensitive to or overdiagnose MDD.

The results of this study suggest that, for the PHQ-9, not all symptoms fit a single depression factor among persons with SCI. Instead, somatic symptoms, especially trouble sleeping, feeling tired, poor appetite, trouble concentrating, and psychomotor changes, do not fit on the same factor with NS symptoms. The etiology of these symptoms may be attributable to the primary medical disorder, disability, or to a depressive illness. Because the specific items that load with each of the 2 factors have varied both within the current study and across the existing studies, the PHQ-9 should be used with caution after SCI.

The presence of 2 factors does not in and of itself preclude use of the instrument, but the variation in which the items load with the 2 factors is of concern. This finding was previously observed in the existing studies and was verified in the current study. This may not be a problem with the measure itself but may reflect the effect of SCI on the assessment of depression using other measures as well (this is an empirical question). The factor structure of this measure argues for an alternative diagnostic approach or the "etiologic" approach used in the DSM-IV [26]. In the etiologic approach, symptoms are counted toward a diagnosis of MDD unless they are clearly and fully accounted for by a general medical condition, or in this case the primary and secondary medical conditions associated with SCI. Optimally, a clinician would review the results of the PHQ-9, conduct a diagnostic interview, and judge whether each symptom is caused by medical or disability related factors versus an underlying depressive or other psychiatric disorder.

Another implication is that the PHQ-9 total score may be spuriously inflated by the endorsement of somatic symptoms. If this is the case, the use of a cutoff score to screen for major depression may lack specificity and overidentify individuals as having probable major depression. If it is not used cautiously, screening could result in more people receiving a diagnosis they may find stigmatizing and may result in wasting clinical resources to perform diagnostic assessments on persons who do not have major depression. Reviewing the consistency of the 2 sets of items will help for diagnostic purposes and for developing treatment plans. In terms of assessment, if the goal is to use the most streamlined measure possible, then the results indicate the 6-item measure would provide reasonable information on both somatic and NS dimensions. However, if the goal is to generate a potential diagnosis, then the clinician should administer the full measure to be consistent with diagnostic procedures. In this case, our results would indicate the clinician could supplement his or her diagnosis by reviewing the consistency of the 2 factors as indicated by the best fit using all 9 items. Clinicians should be aware that the factor structure does not appear to be highly stable and that, for any given individual, there may be different interpretations of somatic and nonsomatic responses.

Finally, the comparatively poor fit among Hispanic participants raises serious concerns about the validity of the PHQ-9 with this population. We simply do not have enough information about the Hispanic sample to evaluate the extent to which this finding is attributable to a language barrier as opposed to real differences in the structure of depressive symptoms. For instance, we do not know the relative portion of individuals who are first or second generation or for whom English is the primary language. Having this type of information would allow us to compare the relative fit of each model for subgroups of Hispanic participants. Because the information is not available, we can only recommend caution when using the PHQ-9 with Hispanic participants.

Limitations

Although the national data set is often treated as optimal for data analytic purposes, the use of these data have several inherent limitations. First, at present no data from inpatients are available that can be used to test the factor structure, and the current study findings were substantially different from those of the only study to use an inpatient SCI sample [22]. Therefore a gap exists in our knowledge between the factor structure of the PHQ-9 between the inpatient period and the 1-year follow-up period, which may reflect some adjustment processes differentially captured by these items.

Second, no uniform way exists to identify or compare those who did and did not participate in the national database collection. Centers have different criteria for inclusion in the catchment area, and response rates certainly vary greatly between centers. The data are not population-based; rather, participants are identified by who meets clinical criteria. Although these considerations are inherent to all types of studies that use the national database, they must be acknowledged regarding this unknown bias.

Third, because only cross-sectional data were used, the results cannot be used to make generalizations regarding changes in the present symptoms or the consistency of the factor structure for given individuals over time. We can say the factor structure was supported by the CFA across participants who have lived varying amounts of time with SCI (not that the factor structure will be invariant over time). Fourth, although the models had generalizability across most subpopulations, Hispanic participants tended to have lower model fit with several different types of solutions. This finding raises the question of whether this population has a fundamentally different structure of depressive symptoms or whether language issues may be clouding the assessment. Furthermore, an inherent limitation exists in using factor analysis in any study given that it is sample dependent;

differences in factor structures across samples may be attributable, in part, to sample characteristics and not only factor structure.

Finally, with respect to clinical implications and to depression studies among people with SCI, only a small number of validity studies comparing screening measures with the "gold standard" of diagnostic assessments have been performed [32]. This study suggests that the utility of somatic symptoms should receive particular scrutiny in validity research.

Future Research

Because the results of this analysis support a 2-factor structure of the PHQ-9 in persons with SCI, the next step in this line of research is to validate each of these dimensions against other ways of measuring depression. Validation against DSM-IV [26] criteria is limited because the PHQ-9 is determined by the criteria for a major depressive episode. Validation against scales that primarily represent psychologic depressive symptoms may be an effective way to examine the cluster of PHQ-9 items representing psychologic symptoms in this sample. Somatic items could be validated against measures of secondary conditions of SCI as a way to further support the 2-dimensional structure of the PHQ-9 in the context of SCI.

Research is needed to identify the utility of different cutoff scores or alternative scoring approaches when generic depression screening measures like the PHQ-9 are used in populations with disabilities or multiple comorbid medical conditions. The costs and benefits of greater false-positive versus false-negative depression screening results may differ by institution, clinical setting, and base-rates of major depression. Trade-offs surrounding screening practices should be discussed as part of clinical policy decisions.

A related need is for investigation of other measures of depressive symptoms in relation to the structure of symptoms in order to determine whether similar 2-factor structures underlie depressive symptoms or whether this finding is due primarily to the PHQ-9 itself. Exploring and comparing factor structures of various depression measures will substantially add to our understanding of the role of somatic symptoms in the experience and manifestation of depression in the context of SCI. Finally, further research needs to be directed at identification of use of factor structure and other facets of measurement of depressive symptoms in the treatment of depressive syndromes after SCI.

CONCLUSION

A 2-factor model, with distinctive somatic and NS factors, appears to best fit responses to the PHQ-9 after SCI.

REFERENCES

1. Elliott TR, Frank RG. Depression following spinal cord injury. Arch Phys Med Rehabil 1996;77:816-823.
2. Frank RG, Elliott TR, Corcoran JR, Wonderlich SA. Depression after spinal cord injury: Is it necessary? Clin Psychol Rev 1987;7:611-630.
3. Kalpakjian CZ, Bombardier CH, Schomer K, Brown PA, Johnson KL. Measuring depression in persons with spinal cord injury: A systematic review. J Spinal Cord Med 2009;32:6-24.
4. Stewart DE. Physical symptoms of depression: Unmet needs in special populations. J Clin Psych 2003;64(Suppl 7):12-16.
5. Williams JW Jr., Noel PH, Cordes JA, Ramirez G, Pignone M. Is this patient clinically depressed? JAMA 2002;287:1160-1170.
6. Bombardier CH, Richards JS, Krause JS, Tulsky D, Tate DG. Symptoms of major depression in people with spinal cord injury: Implications for screening. Arch Phys Med Rehabil 2004;85:1749-1756.
7. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: Validity of a brief depression severity measure. J Gen Intern Med 2001;16:606-613.
8. Beck AT, Ward CH, Mendelson M, Mock J, Erbaugh J. An inventory for measuring depression. Arch Gen Psychiatry 1961;4:561-571.
9. Steer RA, Ball R, Ranieri WF, Beck AT. Dimensions of the Beck Depression Inventory-II in clinically depressed outpatients. J Clin Psychol 1999;55:117-128.
10. Arnau RC, Meagher MW, Norris MP, Bramson R. Psychometric evaluation of the Beck Depression Inventory-II with primary care medical patients. Health Psychol 2001;20:112-119.
11. American Psychological Association. Diagnostic and Statistical Manual of Mental Disorders. 3rd, revised ed. Washington, DC: American Psychological Association; 1987.
12. Aggen SH, Neale MC, Kendler KS. DSM criteria for major depression: Evaluating symptom patterns using latent-trait item response models. Psychol Med 2005;35:475-487.
13. Huang FY, Chung H, Kroenke K, Delucchi KL, Spitzer RL. Using the Patient Health Questionnaire-9 to measure depression among racially and ethnically diverse primary care patients. J Gen Intern Med 2006; 21:547-552.
14. Lamoureux EL, Tee HW, Pesudovs K, Pallant JF, Keeffe JE, Rees G. Can clinicians use the PHQ-9 to assess depression in people with vision loss? Optom Vis Sci 2009;86:139-145.
15. Dum M, Pickren J, Sobell LC, Sobell MB. Comparing the BDI-II and the PHQ-9 with outpatient substance abusers. Addict Behav 2008;33:381-387.
16. Shafer A. Meta-analysis of the factor structures of four depression questionnaires: Beck, CES-D, Hamilton and Zung. J Clin Psychol 2006;61:123-146.
17. Richardson EJ, Richards JS. Factor structure of the PHQ-9 screen for depression across time since injury among persons with spinal cord injury. Rehabil Psychol 2008;53:243-249.
18. Kalpakjian CZ, Toussaint LL, Albright KJ, Bombardier CH, Krause JK, Tate DG. Patient health Questionnaire-9 in spinal cord injury: an examination of factor structure as related to gender. J Spinal Cord Med 2009;32:147-156.
19. Graves DE, Bombardier CH. Improving the efficiency of screening for major depression in people with spinal cord injury. J Spinal Cord Med 2008;31:177-184.
20. MacCallum RC, Browne M, Cai L. Testing differences between nested covariance structure models: Power analysis and null hypotheses. Psychol Methods 2006;11:19-35.
21. Heck RH. Mplus 1.04 software review. Structural Equation Modeling 2000;7:309-317.
22. Krause JS, Bombardier CH, Carter RE. Assessment of depressive symptoms during inpatient rehabilitation for spinal cord injury: Is there an underlying somatic factor when using the PHQ? Rehabil Psychol 2008;53:513-520.
23. Krause JS, Reed KS, McArdle JJ. Prediction of somatic and non-somatic depressive symptoms between inpatient rehabilitation and follow-up. Spinal Cord 2010;48:239-244.
24. Krause JS, Kemp B, Coker JL. Depression after spinal cord injury: Relation to gender, ethnicity, aging, and socioeconomic indicators. Arch Phys Med Rehabil 2000;81:1099-1109.

25. Kemp B, Krause J, Adkins R. Depression among African Americans, Latinos, and Caucasians with spinal cord injury: An exploratory study. Rehabil Psychol 1999;44:235-247.
26. American Psychological Association. Diagnostic and Statistical Manual of Mental Disorders. 4th ed. Washington, DC: American Psychiatric Association; 1994.
27. Kalpakjian CZ, Toussaint LL, Albright KJ, Bombardier CH, Krause JK, Tate DG. Patient health Questionnaire-9 in spinal cord injury: An examination of factor structure as related to gender. J Spinal Cord Med 2009;32:147-156.
28. Muthen L, Muthen B. Mplus, The Comprehensive Modeling Program for Applied Researchers User's Guide. 4th ed. Los Angeles, CA: Muthen & Muthen; 2006.
29. Browne M, Cudeck R. Alternative ways of assessing model fit. In: Bollen K, Long S, eds. Testing Structural Equation Models. Beverly Hills, CA: Sage; 1993, 136-162.
30. Bentler PM. Comparative fit indexes in structural models. Psychol Bull 1990;107:238-246.
31. Brown TA. Confirmatory Factor Analysis for Applied Research. New York, NY: The Gulliford Press; 2006.
32. Kalpakjian CZ, Bombardier C, Schomer K, Brown P, Johnson K. Measuring depression in persons with spinal cord injury: A systematic review. J Spinal Cord Med 2009;32:6-24.

This CME activity is designated for 1.0 AMA PRA Category 1 Credit™ and can be completed online at me.aapmr.org. Log on to www.me.aapmr.org, go to Lifelong Learning (CME) and select Journal-based CME from the drop down menu. This activity is FREE to AAPM&R members and $25 for non-members.

CME Question

The 2-factor structure of depression screening using the PHQ-9 measure in individuals with SCI suggests:

a. depression characterization be done with a combination of nonsomatic and somatic symptoms.
b. 2-dimensional understanding of depression, with more somatic than nonsomatic symptoms.
c. not all symptoms fit a single depression factor.
d. symptom etiology attributable solely to the depressive illness.

Answer online at me.aapmr.org